Collective Attention and the Dynamics of Group Deals

Mao Ye*
Dept. of CSE
Pennsylvania State University
Pennsylvania, USA
mxy177@cse.psu.edu

Chunyan Wang
Dept. of Applied Physics
Stanford University
California, USA
chunyan@stanford.edu

Christina Aperjis
Social Computing Group
HP Labs
California, USA
christina.aperjis@hp.com

Bernardo A. Huberman
Social Computing Group
HP Labs
California, USA
bernardo.huberman@hp.com

Thomas Sandholm
Social Computing Group
HP Labs
California, USA
thomas.e.sandholm@hp.com

ABSTRACT

We present a study of the group purchasing behavior of daily deal customers. Our goal is to understand the dynamics of purchases to ultimately be able to predict how successful a deal will be. We propose a general dynamic model of collective attention for group buying behavior. In our model, the aggregate number of purchases at a given time comprises two types of processes: random discovery and social propagation. We find that these processes are very clearly separated by an inflection point, which in Groupon typically coincides with the "tipping time," the time after which there are enough purchases to guarantee deal transactions. Using data from Groupon we empirically verify our theoretically inferred purchase growth models.

1. INTRODUCTION

Attracting the attention of potential customers in today’s information rich social media is a challenge. As a result, marketers have been forced to target customers in more sophisticated ways. Location-based (regional) and hyper-location-based (within eye-sight) targeting has turned out to be very effective in terms of improving conversion rates from views to purchases [?]. However, since people are unwilling to share their exact locations out of privacy concerns they need to be given some incentive to reveal their position. The most successful incentive employed to date is daily deals [2] In spite of the success of this strategy it is not fully understood what makes it successful and what kind of social behavior the daily deals sites so effectively tap into and exploit. It is clear that deadlines and social propagation play important roles in addition to location-based targeting.

and an interesting question is how to describe the purchasing pattern more precisely so as to be able to predict the future popularity of a deal.

To answer these questions we use renewal theory to model the random discovery process. After tipping, we use a multiplicative process to describe the social propagation behavior and to explain the purchase growth behavior seen on Groupon, the current market leader of daily deals in the USA.

Groupon promotes deals for different geographic markets, or cities, called divisions. In each division, there is typically one featured daily deal. A deal is a coupon for some product or service at a substantial discount off the regular price. Deals may be available for one or more days. Coupons are only redeemable if a certain minimum number of customers purchase the deal, and this number constitutes what they call a "tipping point." Furthermore, sellers may set a maximum threshold size to limit the number of coupons that can be purchased.

Groupon has achieved notable success with its emphasis on high quality local deals, as well as its viral marketing methods. A closer examination of the mechanisms driving user behavior in Groupon could hence provide useful guidance for local marketing campaigns. In this paper we study the evolution of collective attention measured as deal purchases. We base our analysis on data collected from Groupon over a month. Our assumption is that successful deals arise from two behavioral processes: random discovery; resulting from the serendipitous discovery of a deal on the Groupon web portal, or in the mobile app, or via an email subscription; and social propagation; which results from the propagation of deals over social networks. These processes are separated by an inflection point, which in Groupon is the tipping point, after which there are enough purchases to guarantee deal transactions. Before the tipping point is reached the customer base is small so the random discovery process dominates. Conversely, after the tipping point a critical mass of customers have discovered the deal to make social propagation dominate the purchasing behavior.

The contributions of this paper fall into two categories:

• Structure of purchasing dynamics. We present stochastic models that analytically explain the observed

### Permission

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2011 ACM 978-1-4503-0757-4/11/07 ...$10.00.

http://www.bynd.com/2011/05/04/social-loco-research/

http://groupon.com
purchasing behavior.

- Empirical verification of purchase growth and waiting time distribution. We measure the fit of our model to empirical data.

The paper is structured as follows. In Section 2 we discuss related work. In Section 3 we discuss the data sets and the collection strategies used in our study. Section 4 describes our stochastic model and its empirical verification. Section 5 concludes with possible applications of our work and future directions.

2. RELATED WORK

The related work falls into two broad categories, social purchasing behavior, and collective attention.

2.1 Social Purchasing Behavior

According to [5, 7], a buyer’s social network strongly influences her purchasing behavior. In [7], Guo et. al. analyze data from the e-commerce site Taobao to understand how individual’s commercial transactions are embedded in their social graph. In the study, they empirically verify that implicit information passing is present in the Taobao network, and show that communication between buyers is a fundamental driver of purchasing activities. However, according to the study presented in [8], social factors may impose a different level of impact on the user purchase behavior for different e-commerce products.

Several studies have been conducted to understand the success of Groupon. In [1], Arabshahi examined the business model of Groupon. In [9], Utpal conducted a survey-based study on Groupon, in order to understand how businesses fare when running such promotions. In [8], Edelman et al. consider the benefits and drawbacks of using Groupon from the merchant’s point of view, modeling whether advertising and price discrimination effects can make group discounts profitable. The paper that is most related to our work is [3], where data analysis on the purchase history of Groupon deals have been conducted. One key outcome of [3] is the preliminary evidence that Groupon is behaving strategically to optimize deal offerings, giving customers “soft” incentives (e.g., deal scheduling and duration, deal featuring, and limited inventory) to purchase. Our work differs from these studies by focusing on modeling the deal purchasing dynamics over time and by highlighting the importance of the tipping point and its implication to social propagation.

2.2 Collective Attention

Recent studies of collective attention on social media sites such as Twitter, Digg and YouTube [10, 9, 2] have clarified the interplay between popularity and novelty of user generated content. The allocation of attention across items was found to be universally log-normal, as a result of a multiplicative process that can be explained by an information propagation mechanism inherent in all these sites. While the specific time scales over which novelty decays differ between different systems depending on their typical type of content, the functional form of the decay is predictable.

3. DATASET

We collected data from Groupon’s socially promoted and local daily deal websites in the US. Groupon also provides a convenient API which allows us to obtain more detailed information about the deals.

By the end of April 2011, Groupon’s business covered about 120 cities in the US. We monitored all Groupon deals offered in 60 different randomly selected cities during the period between May 4th and June 16th, 2011. In total we collected the entire purchase traces of 4349 deals.

Each deal is associated with a basic set of features: the deal description, the retail and discounted prices, the start and end dates, the “tipping points” required for the deal to be “on”, the number of coupons sold, whether the deal was available in limited quantities, and if it sold out. We monitored the number of purchases and the position of each deal in 20-minute time intervals. A surprisingly large portion, 41%, of all deals exhibited non-monotonically increasing behavior. We assume that this indicates that something was wrong with the deal, e.g. false marketing due to an inflated list price, and customers who initially purchased the deal requested a refund (an option Groupon supports and markets). Due to the unknown user behavior behind these deals we exclude these deals from our initial study. We, however, intend to look more closely at this phenomena in future work. Hence, 2563 deals were left to analyze.

In our dataset, 190 deals (out of 2563) had not reached their tipping point when they expired. In the following, these deals are called failed deals; and deals that are turned on successfully are called tipped deals.

4. PURCHASE DYNAMICS

In this section, we introduce a model of the purchase dynamics of Groupon deals. As previously mentioned, a Groupon deal is generally discovered by the user in one of the following four ways: (1) by visiting the Groupon webpage, (2) by running a smart-phone application, (3) by getting notifications via email and (4) by communicating with friends. The first three are referred to as random discovery and the fourth is referred to as social propagation.

Based on this observation, our model describes the purchase dynamics as follows. Let $N_t$ denote the number of times that the deal has been purchased at time $t$. We then have

$$N_{t+\Delta t} - N_t = \alpha_t \cdot Y_t + \beta_t \cdot f(t, N_t),$$

where $\alpha_t$ and $\beta_t$ are weight factors, $Y_t$ is a non-negative random variable denoting the number of purchases caused by random discovery in the interval $(t, t+\Delta t)$, and $f(t, N_t)$ represents the count of purchases caused by social propagation in the same interval as a function of $t$ and $N_t$.

Figure 1 shows how the number of purchases increases over time when the tipping point is equal to 10 (the most frequent value) and 20 purchases. The graph is based on 492 (resp. 477) deals whose tipping point was equal to 10 (resp. 20) in our dataset. We observe the same pattern for deals with other tipping points, e.g., 5 and 30. We find an approximately linear growth of purchases at the beginning of the lifetime of a deal. For both tipping points, the purchase rate is relatively small and steady before the tipping time. After tipping or around tipping, the purchases grow.

\[\text{Taobao is a Chinese Consumer Market place, and also the world’s largest e-commerce website, http://www.taobao.com.}\]

\[\text{http://www.groupon.com/pages/api}\]
dramatically for about 11.6 hours, after which the purchase rate decreases. The tipping time, thus, typically coincides with the inflection point in the purchase dynamics.

Note that the number of purchases of a deal with a tipping point of 10 purchases is usually smaller than the corresponding number of a deal with a tipping point of 20, even though we do not observe a significant difference before the tipping times. One possible reason is that deals tipping after 10 purchases have smaller purchase populations than those that tip after 20 purchases, depending on the specific categories of products and services. Furthermore, the potential purchase population may also act as the reference for Groupon and local merchants when they set the tipping point for a deal.

Based on these findings we write our equation as:

\[ N_{t+\Delta t} - N_t = \begin{cases} Y_t & \text{before tipping} \\ r(t)X_tN_t & \text{after tipping} \end{cases} \]  

Thus, we are implicitly assuming that before tipping \( \alpha_t = 1 \) and \( \beta_t = 0 \), whereas after tipping \( \alpha_t = 0 \) and \( \beta_t = 1 \) in Figure 1. This assumption is motivated by the fact that random discovery dominates before the deal is tipped and social propagation dominates afterwards — even though the two processes may coexist. In particular, before the tipping point the customer base is small so the random discovery process dominates. Conversely, after the tipping point a critical mass of customers have discovered the deal to make social propagation dominate the purchasing behavior.

According to Equation 2, after a deal is tipped, the increase in the number of purchases \((N_{t+\Delta t} - N_t)\) is proportional to the number of people that has purchased the deal up to time \( t \). Intuitively, a fraction of the people that already purchased the deal will notify some of their friends about it, and a fraction of these friends will purchase the deal. These fractions are represented by the positive random variable \( X_t \). We assume that \( \{X_t\} \) are independent and identically distributed random variables. Since \( X_t \) is assumed to be positive, \( N_t \) can only increase over time. This growth in time is eventually curtailed by a decay in novelty, which is parameterized by the factor \( r(t) \). As we discuss later, \( r(t) \) is decreasing in \( t \).

### 4.1 Purchase Dynamics Before Tipping

We denote by \( \tau_t \) the interarrival times of purchases. In particular, \( \tau_t \) is the time between the \( i-1 \) and the \( i \)-th purchases. Suppose that each \( \tau_t \) is independently drawn from some distribution \( F \). We denote a deal’s tipping point by \( \theta \), that is, the number of purchases required to turn on the deal. Let \( L \) be the total time that the deal is open for purchases (as set by the seller). Then, \( N_L \) is the final number of purchases when the deal ends.

Let \( F_\theta \) denote the \( n \)-fold convolution of \( F \). Then, \( F_\theta \) is the distribution of the sum of \( n \) consecutive interarrival times. Thus, the distribution of the tipping time for deals with the same tipping point \( \theta \) is given by \( F_\theta \), the \( \theta \)-fold convolution.

We now look at the probability that a deal fails, i.e. does not turn on (tip). According to Groupon’s mechanism, a deal is said to be turned on as long as its number of purchases reaches the tipping point \( \theta \) before the deal expires, i.e. its lifetime \( L \) ends. So the probability of a deal failing is equal to \( \Pr(N_L < \theta) \).

\[
\Pr(N_L < \theta) = \sum_{n=1}^{\theta-1} \Pr(N_L = n) \quad (3)
\]

Since the \( \tau_t \) variables are iid interarrival times of purchases, it follows that this is a renewal process. We use \( S_n = \sum_{i=1}^{n} \tau_i \) to denote the time spent until the \( n \)-th purchase.

It is easy to see that \( N_t = \text{sup}\{n : S_n \leq t\} \), and thus,

\[
\Pr(N_t = n) = \Pr(N_t \geq n) - \Pr(N_t \geq n+1) = \Pr(S_n \leq t) - \Pr(S_{n+1} \leq t) = F(t) - F_{n+1}(t) \quad (4)
\]

Applying Equation 4 to Equation 3, we have:

\[
\Pr(N_L < \theta) = \sum_{n=1}^{\theta-1} (F(n) - F_{n+1}(n)) = (F(L) - F_{\theta-1}(L)) \quad (5)
\]

Note that Equation 5 can predict the failure ratio (i.e., the probability not to be turned on) of a deal. Conversely, using this equation, given the failure ratio, we can estimate the parameters of \( F \), such as the mean value.

This analytical model can be easily extended to predict the probability that a deal will be turned on when we know the number of purchases up to a given point in time. For example, if at time \( t_1 \), a deal has already got \( n_1 \) purchases, then the probability that the deal will be turned on can be estimated as

\[
\Pr(N_L < \theta | N_{t_1} = n_1) = F(L - t_1) - F_{\theta-n_1}(L - t_1) \quad (6)
\]

We now consider what distribution the interarrival times follow in Groupon. To exclude the impact of tipping point differences, we first consider only deals with a tipping point of 10 purchases (the tipping point distribution mode) from all the data we gathered. As shown in Figure 2, interarrival times follow an exponential distribution. Thus, before tipping, the arrival rate of purchases follows a Poisson process.

This observation confirms our assumption about random discovery, since if a user randomly checks the websites or a
smartphone app the probability of a purchase taking place in the next infinitely-small time interval is the same, and hence the intervals between purchases follow an exponential distribution.

### 4.2 Purchase Dynamics After Tipping

We now focus on the dynamics after tipping, and for expositional clarity consider a deal’s tipping point as time 0. Thus, $N_0$ denotes the tipping point of a deal (i.e., the number of purchases of a deal when it was tipped). Then, according to Equation (2), the number of purchases at time $T$ (that is, $T$ time units after the deal tipped) is given by

$$N_T = \prod_{t=1}^{T} (1 + r(t)X_t)N_0$$

Note that the realization of $X_t$ will in general be different in different time periods; however all random variables $X_t$ follow the same distribution. When $X_t$ is small (which is the case for small time steps), we have the following approximate solution for $N_T$:

$$N_T \approx \prod_{t=1}^{T} e^{r(t)X_t}N_0 = e^{\sum_{t=1}^{T} r(t)X_t}N_0.$$  

Taking the logarithm on both sides, we get

$$\log N_T - \log N_0 \approx \sum_{t=1}^{T} r(t)X_t.$$  

The decay factor $r(t)$ is estimated according to Equation 2 and Equation 3 as follows:

$$r(t) = \frac{E(\log N_t) - E(\log N_{t-1})}{E(\log N_t) - E(\log N_0)}.$$  

where we normalize $r(1)$ to 1.

In Figure 3 we plot the novelty decay $r(t)$ for the first 16 hours after tipping, as estimated from our dataset. Recall that in this section $N_0$ denotes the tipping point, and time $t = 0$ is the tipping time. We observe that $r(t)$ decreases over time. Moreover, Figure 3(b), suggests that the novelty decay is exponential during the first 16 hours after tipping. In particular,

$$r(t) \approx \exp(at + b),$$

where $a = -0.064$ and $b = -0.29$. The $R^2$ value for this fit is 0.8839.

Here, we are interested in evaluating how well our model helps explain the purchase growth after a deal has tipped. With both $a$, $b$ estimated, we can use our results to explain the growth of purchases. In Figure 4 we demonstrate the potential predictive power of our model by empirically verifying the growth of purchases of deals after they have tipped. We conducted a time-series cross correlation test over the derived and empirical curves, and get an $R^2$ value of 0.9404.

### 5. CONCLUSIONS
In this paper, we studied the collective attention and purchase dynamics of group deals. We used Groupon as a case study to empirically verify our models. While we concentrated on data from Groupon, we stress that our model is general enough so as to capture the purchase behaviors in group deals with both random discovery and social propagation behaviors. In future work we intend to extend our model to capture purchase cancelation behavior, multi-day dynamics, as well as decay factor dependency on tipping point. Furthermore, we plan to conduct baseline, and state-of-the-art benchmark prediction tests of a predictor of final purchase volume derived from our model. We are also interested in studying additional group deal sites to see how social propagation mixes with random discovery in settings where there may not be any clear inflection point such as the tipping point in Groupon.

6. REFERENCES